HYPOTHESIS TESTING

- In hypothesis testing, evaluates 2 or more exclusive statement on a Population using sample data.

NULL HYPOTHESIS → It is a hypothesis that says there is no statistical significance between two variable in the hypothesis.

Alternate hypothesis → There is statistically significant relationship between variable. It is inverse/opposite of Null hypothesis

Steps in hypothesis testing –

1) Make Initial Assumption (H_0). H_0 → Null hypothesis . H_1 → Alternate hypothesis
2) Collect Data. (evidences)
3) Gather evidences to reject the Null hypothesis or not reject null hypothesis.

This lead us to Confusion Matrix –

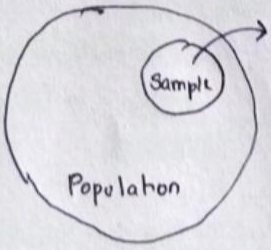
	H_0	H_1
Accept / Do not refuse	OK	Type 2 Error
Reject	Type 1 Error.	OK

Type 1 error – Due to lack of evidence we have to reject NULL hypothesis/H_0.
For example, we all know Salman Khan killed street side vendor by his car, but due to lack of evidence we have to reject Salman Khan is guilty. But due to this Type 1 error, we lost the justice to the life of people killed in the crash.

Type 2 error – When we reject the alterate hypothesis and accept the null hypothesis. even if null hypothesis is wrong.

Type 1 error, in which you mistakenly conclude an effect is real, when it is really just due to chance.
Type 2 error, in which you mistakenly conclude than an effect is not real, (due to chance), when it really is real

H_0 → Null hypothesis , H_1 → Null hypothesis , α – Significance level



→ Statistic.

$P(\text{Statistic} \mid H_0 \text{ true})$ → p-value.

p-value $< \alpha$ → reject H_0 , p-value $\geq \alpha$ → Fail to reject H_0

Reality check →

	Reality: H_0 true	Reality: H_0 false
reject H_0	Type 1.	correct conclusion
Fail to reject H_0	correct conclusion	Type 2.

Type 1 error because in reality H_0 is true but we reject it. It cannot happen.
Type 2 error. H_0 is false but we failed to reject H_0. This is partiality.

Example – Consider a website. Suppose it is of white background and avg time spent by customer is equal or less than 20 minutes. Someone said if we change the colour, mean avg spend time will be more than 20 minutes.

Step 1 – H_0: $\mu \geq 20$ minute after change H_1: $\mu < 20$ minute after change.

Step 2 – Significance level : $\alpha = 0.05$

Step 3 – Take sample $n = 100$. Suppose sample mean $\bar{x} = 25$ minutes.

Step 4 – p value : $P(\bar{x} \geq 25 \text{ minutes} \mid H_0 \text{ true})$

Step 5 – pvalue $< \alpha \Rightarrow$ Reject H_0, suppose Pvalue $= 0.03$.
pvalue $> \alpha \Rightarrow$ Do not reject H_0, suppose Avalue $= 0.5$.